Backup System and Method

Field of the invention

[0001] This invention generally relates to computer systems and more particularly to backup systems and methods.

Background and prior art

[0002] Most computer users in the desktop environment employ some sort of data protection procedures and devices to protect computer system files and data files in the event of an accidental delete or overwrite of these files from the primary storage medium, which is usually the computer's hard drive. Computer data protection systems usually involve the backup of system and data files onto some sort of secondary storage device utilizing removable storage media, such as floppy disk drives, other hard disk drives, tape drives, etc. More sophisticated users may utilize automatic data protection devices and procedures that backup the entire system on a regular basis, allowing for a full system recovery if needed.

[0003] For example, tape backup systems are used to protect data files and other information from computer system failures such as hard disk crashes or computer virus attacks. A tape backup system stores this data on removable off-line media (i.e., the tape); this data can then be retrieved in the event of data loss. Conventional backup systems typically provide the end user with a choice of making a "full" backup or an "intermediate" or "modified" backup. Full backups make complete copies of all the data on the computer to a set of one or more backup tapes. Incremental backups are generally much smaller than full backups since they simply save the data that has been changed since either the last full backup or the most recent incremental backup. Example of such prior art backup systems are described in U.S. Pat. No. 5,276,860, U.S. Pat. No. 5,758,067, U.S. Pat. No. 6,212,512, and U.S. Pat. No. 6,330,570.

[0004] Further, a backup system for Enterprise Data Protection and Disaster Recovery is commercially available from Hewlett Packard, i.e. HP OpenView Storage Data Protector.

Summary of the invention

[0005] The present invention provides a backup system for storing of data objects on secondary storage devices. The backup system has a number of buffer memories which provide interfaces to the secondary storage devices. Preferably, the backup system is configurable to couple at least a sub-set of the buffer memories in one or more daisy-chains.

[0006] This way a user is enabled to control the copying or mirroring of data objects with a fine level of granularity. Instead of mirroring all backup data, the invention enables to mirror only a sub-set of the data objects of the backup data. This way mirroring can be limited to the most important data objects. Likewise, in the case of copying the invention enables to copy only a sub-set of the data objects which is configurable by the user. This way a high degree of flexibility is provided which can be used in order to improve tape management.

[0007] In particular the present invention can be advantageously used for tape compaction, tape refreshing, tape demultiplexing, restore chain optimisation and data object migration.

[0008] In accordance with a further preferred embodiment of the invention at least one backup media agent is used for the purposes of backup, mirroring and/or copying. A backup media agent comprises a plurality of buffer memories and a program module for writing of data from the buffer memories to a secondary storage device being assigned to the backup media agent.

[0009] Usually there is a one-to-one relationship between backup media agents and secondary storage devices. Preferably each one of the buffer memories of the backup media agent is assigned to one data object. The

program module of the backup media agent combines the data sequences of the individual data objects into a single data sequence for storage on the secondary storage device.

[0010] In accordance with a further preferred embodiment of the invention backup media agents are cascaded in order to provide one or more mirroring stages.

[0011] In accordance with a further preferred embodiment of the invention at least one restore media agent is used for reading of data from the secondary storage device. The restore media agent has a plurality of buffer memories and a program module for reading of data sequences of data objects stored on the secondary storage device. The program module splits the single data sequence received from the secondary storage device into separate data objects and stores the data objects in separate ones of the buffer memories from where the data objects are forwarded to other buffer memories of backup media agents.

[0012] In accordance with a further preferred embodiment of the invention a number of client computers is coupled to a backup server. The client computers have local backup agents for reading of data objects to be protected from a primary data source of the respective client and for assigning of an unequivocal identifier to each data object. The unequivocal identifier assigned to the data objects enables the backup media agent to combine a plurality of data object sequences into a single data sequence for sequential storage on the secondary storage device.

[0013] It is to be noted that instead of tape recorders any other suitable storage technology which provides removable storage media can be used.

Brief description of the drawings

[0014] In the following preferred embodiments of the invention will be described, by way of example, and with reference to the drawings in which:

[0015] Figures 1A and 1B together is a block diagram of a client-server backup system,

[0016] Figures 2A and 2B are illustrative of the assignment of data sources to backup media agents and of a daisy-chain configuration of the buffer memories of the backup media agents,

[0017] Figure 3 is illustrative of a flow chart for performing a backup and data mirroring procedure,

[0018] Figure 4 is a block diagram of a backup server being configured to perform a copying procedure,

[0019] Figures 5A and 5B are illustrative of the coupling of buffers of restore media agents and backup media agents for the purposes of data copying.

Detailed description

[0020] Figures 1A and 2A together is a block diagram of backup system 100. Backup system 100 has backup server 102 to which a plurality of client computers 104, 106, 108, ... are connected by means of data bus 110. For example data bus 110 is an Ethernet.

[0021] Client computer 104 has primary storage device 112 for local storage of client data. Typically primary storage device 112 is a disk mass storage device. Various files and folders are typically stored on primary storage device 112. ID generator 116 generates an unequivocal identifier for the data sequence. Each data packet in the sequence is labelled with that unequivocal identifier in order to be able to identify a given data packet as belonging to a

particular data sequence at a later point of time, such as for the purpose of performing a restore operation.

[0022] Each one of the files and folders stored on primary storage device 112 can be selected by a user for data protection, i.e. for backing up the selected file or folder at regular prescheduled or user initiated points of time. The user selected files or folders stored on primary storage device 112 provide a data source D1 during a backup procedure.

[0023] Further client computer 104 has data source backup agent 114 which reads the data to be backed up from primary storage device 112. Data source backup agent 114 creates a data sequence containing the data read from primary storage device 112. Data source backup agent 114 is realised as a program module of a tape backup software package. Client computers 106, 108, ... have an equivalent design as client computer 104. Client computer 106 has data source D2; client computer 108 has data source D3, ...

[0024] Backup server 102 has program module 118 for providing a user interface. By means of the user interface a user is enabled to configure backup server 102 for the purposes of backup, mirroring and/or copying. The corresponding configuration data is stored in configuration file 120. In configuration file 120 the data sources D1, D2, D3, ... to be backed up are specified. Further configuration file 120 enables to define one or more backup groups. In the example considered here one backup group 122 has been defined. Backup group 122 has a number of backup media agents 124, 126, ...

[0025] Backup media agent 124 has a number of buffer memories 128 and a program module 130. Program module 130 belongs to the tape backup software package which includes data source backup agents 114 installed on the client computers. Program module 130 is designed to interact with data source backup agent 114 in order to receive backup data from the respective client computer on which data source backup agent 114 is installed.

[0026] Buffer memories 128 of backup media agent 124 provide intermediate storage for the backup data received from one of the data source backup agents 114 for the purpose of buffering the respective backup data stream flowing from the respective client computer to backup server 102. Typically buffer memories 128 are random access memories or first-in first-out (FIFO) buffers. Backup media agent 124 is assigned to one of the tape recorders 132 being coupled to backup server 102. Each one of the buffer memories 128 of backup media agent 124 is assigned to one of the data sources D1, D2, D3, ... to be backed up.

[0027] For example the backup data sequence provided by data source backup agent 114 of client computer 104 for data source D1 is buffered in one of the buffer memories 128. The individual data sequences delivered by the data source backup agents 114 which are buffered in buffer memories 128 are combined by program module 130 into a single data sequence which is sequentially stored by means of one of the tape recorders 132 which is assigned to backup media agent 124.

[0028] The further backup media agents 126, ... of backup group 122 are of similar design as backup media agent 124. These backup media agents 126, ... are assigned to other data sources and to other recorders than backup media agent 124.

[0029] Further a user can configure one or more mirror groups 134, 136, ... for providing one or more data mirroring stages. In the example considered here two data mirroring stages, i.e. mirror group 1 and mirror group 2, have been configured. Mirror group 1 has backup media agents 138, 140, ... Backup media agents 138, 140, ... are of similar design as backup media agents 124, 126, ... of backup group 122. Backup media agents 138, 140, ... of mirror group 1 are coupled to backup media agents 124, 126, ... of backup group 122 for "mirroring" of the backup data which is written on tape by backup media agents 124, 126, ... of backup group 122. "Mirroring" means that an additional copy of the backed up data is created by writing the data to another

tape by means of another tape recorder for improved data security. By means of mirroring a user can obtain multiple copies of the backup data on different tapes which can be stored at different locations.

[0030] Backup media agents 138, 140, ... are configured differently, for the purpose of data mirroring. Buffer memories of backup media agents 138, 140, ... are assigned to buffer memories of backup media agents 128, 126, ... Each one of backup media agents 138, 140, ... of mirror group 1 is assigned to one of the recorders 132 which are coupled to backup server 102 for sequential writing of mirrored backup data.

[0031] Mirror group 2 is of a similar structure as mirror group 1. The buffer memories of the backup media agents of mirror group 2 are assigned to buffer memories of backup media agents of mirror group 1 in order to provide a second data mirroring stage.

[0032] Further backup server 102 has control program 142 which controls the backup and/or mirroring and/or copying process as specified in configuration file 120.

[0033] It is to be noted that the various backup media agents as well as user interface 118 can be implemented in a distributed way on multiple servers.

[0034] Figures 2A and 2B are illustrative of a user defined configuration of backup server 102 of figure 1. In the example considered here backup media agent 124 of backup group 122 is coupled to tape recorder R1; backup media agent 126 is coupled to tape recorder R2; backup media agent 138 of mirror group 1 is coupled to tape recorder R5; backup media agent 140 of mirror group 1 is coupled to tape recorder R8; backup media agent 144 of mirror group 2 is coupled to tape recorder R3 and backup media agent 146 of mirror group 2 is coupled to tape recorder R6. Preferably the assignment of the backup media agents to tape recorders is also stored in configuration file 120 (cf. figure 1).

[0035] Buffer memory B1 of backup media agent 124 is assigned to data source D3; buffer memory B2 of backup media agent 124 is assigned to data source D2, etc. In other words, at least a sub-set of the buffer memories of backup media agents 124, 126, ... of backup group 122 is assigned to at least a sub-set of the data sources D1, D2, D3, ... Preferably the assignment of data sources to backup media agents and/or to individual buffer memories of the backup media agents of backup group 122 is stored in configuration file 120.

[0036] As far as mirror group 1 is concerned buffer memory B2 of backup media agent 138 is coupled to buffer memory B2 of backup media agent 124; buffer memory B3 of backup media agent 138 is coupled to buffer memory B3 of backup media agent 124, etc. In other words at least a sub-set of the buffer memories of backup media agents 138, 140, ... of mirror group 1 is coupled to at least a sub-set of the buffer memories of backup media agents 124, 126, ... of backup group 122. Likewise a sub-set of the buffer memories of backup media agents 144, 146, ... of mirror group 2 is coupled to a sub-set of the buffer memories of backup media agents 138, 140, ... of mirror group 1.

[0037] Again the assignment of buffer memories of mirror group to buffer memories of the backup group 122 is user configurable and can be stored in configuration file 120 as well as the assignment of buffer memories of mirror group 2 to buffers of mirror group 1.

[0038] In operation the backup media agents 124, 126, ... of backup group 122 receive backup data sequences from the respective data sources D1, D2, ... The data sequences are buffered and combined into single data sequences by programs 130 for sequential storage on the respective tape recorders. Further the data sequences received from the individual data sources are forwarded through the respective buffers to the corresponding buffers of mirror group 1. Here the individual data sequences are also combined into single data sequences for sequential storage on the respective tape recorders. The same applies analogously to mirror group 2.

[0039] The cascaded connection of buffers of one or more mirror groups to the backup group 122 results in a daisy-chain configuration of the buffer memories. For example a daisy-chain is created which comprises buffer memory B2 of backup media agent 124, buffer memory B2 of backup media agent 138 and buffer memory B2 of backup media agent 144 as apparent from figure 2. Another example for a cascaded connection of buffers creating a daisy-chain is buffer memory B3 of backup media agent 126 coupled to buffer memory B1 of backup media agent 140 which in turn is coupled to buffer memory B1 of backup media agent 146.

[0040] By configuring the daisy-chains the user can control the copying or mirroring of data objects with a fine level of granularity. This is due to the fact that the selection of backup media agents of the backup group and one or more of the mirroring groups implies the user's selection for creating one or more additional backup copies for the data objects which are backed-up by the back up media agents of the backup group. Usually each one of the data sources D1, D2, ... provides data for creating backup copies of data objects of the respective data source. Usually data objects which are backed-up are individual files, folders or a directory or sub-directory containing multiple files and/or folders.

[0041] It is to be noted that in case backup media agents run on the same server a buffer memory can be shared between two backup media agents in order to prevent unnecessary copying. For example, backup media agents 126 and 140 can use a single buffer memory 128 with three buffers B1, B2 and B3 instead of six buffers as shown in the example of Figs. 2A and 2B.

[0042] Figure 3 shows a flow chart illustrating a backup and mirroring process. In step 300 a data object is read from a primary storage device, such as a disk, on one of the client computers by the data source backup agent and an unequivocal identifier is assigned to the data packets of the resulting data sequence. The data sequence of the data object is transmitted to the backup server in step 302.

[0043] In step 304 the data sequence of the data object is stored in a buffer of the backup media agent (BMA) which is assigned to the data object to be backed up. In step 306 the data stream of the object is written to a tape by means of a tape recorder being assigned to the backup media agent. The backup media agent may receive one or more additional data sequences from other data objects. If this is the case these individual data sequences of the data objects are combined into a single data sequence which is stored on the tape.

[0044] In step 308 the data sequence of the object is forwarded from the buffer memory to backup media agent of the first level of mirroring in step 308. In step 310 the data sequence is stored by the backup media agent on the first level of data mirroring on a tape by means of the tape recorder being assigned to that backup media agent. Again the backup media agent of the first data mirroring stage may receive other data sequences from other buffer memories of the backup group. In this case the individual data sequences are combined into a single data sequence which is written from the backup media agent of the mirror group 1 to its assigned tape recorder. This is done in step 312.

[0045] In step 314 the data sequence of the data object to be backed up is forwarded from the buffer of the backup media agent of mirror group 1 to a buffer of backup media agent of mirror group 2 for storage on a tape. This process can go on in order to provide multiple further mirroring stages.

[0046] Figure 4 shows backup server 102 in a copying configuration. The copying configuration is used for selectively copying backed up data from previously recorded backup tapes. For the purpose of copying the data needs to be read from the backup tapes which is done by means of restore media agents. Restore media agents have buffer memories and program module 158 of the tape backup software package. Usually restore media agents are activated for recovery of backup data, such as for the purpose of migration or disaster recovery. However in the preferred embodiment of figure 4 the restore

media agent are used for the purpose of selective copying of previously backed-up data.

[0047] Backup server 102 has restore group 148 which comprises a number of restore media agents 150, 152, ... Further backup server 102 has copy groups 154, 156, ... for providing one or more copying stages. In the example considered here two copying stages have been configured. Copy group 1 has backup media agent 164, 166, ...

[0048] The configuration of restore group 148 and copy groups 1 and 2 as well as the assignment of data sources to the restore media agents 150, 152 to data sources and recorders as well as the configuration of the copy groups and the assignment of the respective backup media agents to tape recorders 132 is stored in configuration file 120 and can be entered by means of program module 118 which provides the user interface.

[0049] Restore media agent 150 provides for an inversion of a backup operation performed by a backup media agent. For this purpose restore media agents 150, 152, ... have program modules 158. For example program module 158 of restore media agent 150 reads a single data sequence from the recorder 132 which is assigned to restore media agent 150 and splits the single data sequence into data sequences of individual data objects. The individual data sequences are stored in buffer memories 128 from where they are forwarded to buffer memories 128 of copy group 1. From there the individual data sequences are forwarded to buffer memories of copy group 2 to provide two stages of copying.

[0050] Figures 5A and 5B show in more detail a configuration of the restore and copy groups by way of example.

[0051] Restore media agent 150 is coupled to tape recorder R2 whereas restore media agent 152 is coupled to tape recorder R5. Backup media agent 158 of copy group 1 is coupled to tape recorder R1 and backup media agent 160 is coupled to tape recorder R3. Backup media agent 164 of

copy group 2 is coupled to tape recorder R6 and backup media agent 166 of copy group is coupled to tape recorder R8.

[0052] Buffer memories B1 of restore media agent 150, B1 of backup media 158 and B2 of backup media agent 164 are connected in a daisy-chain. Further buffer memories B2 of restore media agent 150, B1 of backup media agent 160 and B3 of backup media agent 166 are connected in a daisy-chain configuration, etc.

[0053] In operation programs 158 of restore media agents 150, 152, ... read respective single data sequences from tape recorders R2 and R5. By means of the unequivocal identifiers assigned to the data packets of the single data sequences the single data sequences are split into individual data sequences of individual ones of the backed up data objects and stored in respective ones of the buffer memories 128. From there the individual data sequences are forwarded to buffers of copy group 1 and copy group 2 where respective copies are written on tape by means of the respective tape recorders assigned to the backup media agents of the copy groups.